AD_____

Award Number: W81XWH-12-1-0013

TITLE: Whole Exome Analysis of Early Onset Alzheimer's Disease

PRINCIPAL INVESTIGATOR: Margaret Pericak-Vance

CONTRACTING ORGANIZATION: University of Miami, Miami, FL 33136

REPORT DATE: April 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
April 2014	Annual Report	1 March 2013 to 31 March 2014

4. TITLE AND SUBTITLE
Whole Exome Analysis of Early Onset Alzheimer's Disease

5a. CONTRACT NUMBER
W81XWH-12-1-0013

5b. GRANT NUMBER
W81XWH-12-1-0013

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Margaret A. Pericak-Vance, Ph.D.

E-Mail: mpericak@med.miami.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Miami, 1400 NW 10th Avenue, Miami, FL 33136

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The primary focus toward identification of Alzheimer disease (AD) risk genes over the past five years has been testing the common disease common variant (CDCV) hypothesis through the use of genome-wide association studies (GWAS) in late onset Alzheimer disease (LOAD). While common variation clearly plays a role in AD there is a growing realization that the CDCV hypothesis is unlikely to explain all the genetic effect underlying AD. One alternative hypothesis invokes multiple rare variants (RV) in one or more genes, each with stronger individual effects than CDCV genes. We designed this project to test the rare variant hypothesis in AD by examining those cases with the most severe phenotype as determine by early onset (EOAD, cases with AAO < 60 years). Although there are three known EOAD genes (PS1, PS2 and APP) they account for only ~60-70% of familial EOAD and even less of sporadic EOAD. Thus, the majority of the genetics of EOAD remains unknown. Until now, large extended families with AD in multiple generations were necessary to identify variants of significant effect contributing to AD risk, however, with the advent of new genomic technologies such as high-throughput sequencing technology, small family aggregates and isolated cases, particularly those with an extreme phenotype of the disorder (such as early onset) can be used. Thus, we will utilize whole exome high-throughput sequencing to identify high risk AD variants that we will further characterize with respect to AD. We will examine both Caucasian and Caribbean Hispanic AD populations. Our two pronged approach includes structural characterization at the DNA level (Dr. Pericak-Vance), and analysis of Caribbean Hispanics (Dr. Richard Mayeux). Comparing across populations will be extremely useful. Specifically, high priority RVs identified through the whole exome analysis will be further explored with multiple strategies. We will also genotype the interesting variants in a large sample of late-onset (LOAD) cases to examine their involvement in all AD. We will thus prepare a list of high priority candidates for additional follow-up and functional analysis.

.

15. SUBJECT TERMS
none provided

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	30	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

INTRODUCTION:

The primary focus toward identification of Alzheimer disease (AD) risk genes over the past five years has been testing the common disease common variant (CDCV) hypothesis through the use of genome-wide association studies (GWAS) in late onset Alzheimer disease (LOAD). While common variation clearly plays a role in AD there is a growing realization that the CDCV hypothesis is unlikely to explain all the genetic effect underlying AD. One alternative hypothesis invokes multiple rare variants (RV) in one or more genes, each with stronger individual effects than CDCV genes. We designed this project to test the rare variant hypothesis in AD by examining those cases with the most severe phenotype as determine by early onset (EOAD, cases with AAO < 60 years). Although there are three known EOAD genes (PS1, PS2 and APP) they account for only ~60-70% of familial EOAD and even less of sporadic EOAD. Thus, the majority of the genetics of EOAD remains unknown. Until now, large extended families with AD in multiple generations were necessary to identify variants of significant effect contributing to AD risk, however, with the advent of new genomic technologies such as high-throughput sequencing technology, small family aggregates and isolated cases, particularly those with an extreme phenotype of the disorder (such as early onset) can be used. Thus, we will utilize whole exome high-throughput sequencing to identify high risk AD variants that we will further characterize with respect to AD. We will examine both Caucasian and Caribbean Hispanic AD populations. Our two pronged approach includes structural characterization at the DNA level (Dr. Pericak-Vance), and analysis of Caribbean Hispanics (Dr. Richard Mayeux). Comparing across populations will be extremely useful. Specifically, high priority RVs identified through the whole exome analysis will be further explored with multiple strategies. We will also genotype the interesting variants in a large sample of late-onset (LOAD) cases to examine their involvement in all AD. We will thus prepare a list of high priority candidates for additional follow-up and functional analysis.

BODY:

WES and variant prioritization

Whole exome sequencing (WES), quality control and variant calling, variant annotation, and variant filtering is complete on 55 samples submitted by Columbia University to the University of Miami. Additionally, WES and analysis of 51 samples from 46 multiplex families from The University of Miami and Vanderbilt University is complete. Identity-by-descent analysis of Hispanic families was also performed. Following these analyses, comparison of the candidate variants/genes shared across Hispanic families and NH-white cases was done. From these analyses, a list of 125 unique variants was prioritized for follow-up genotyping.

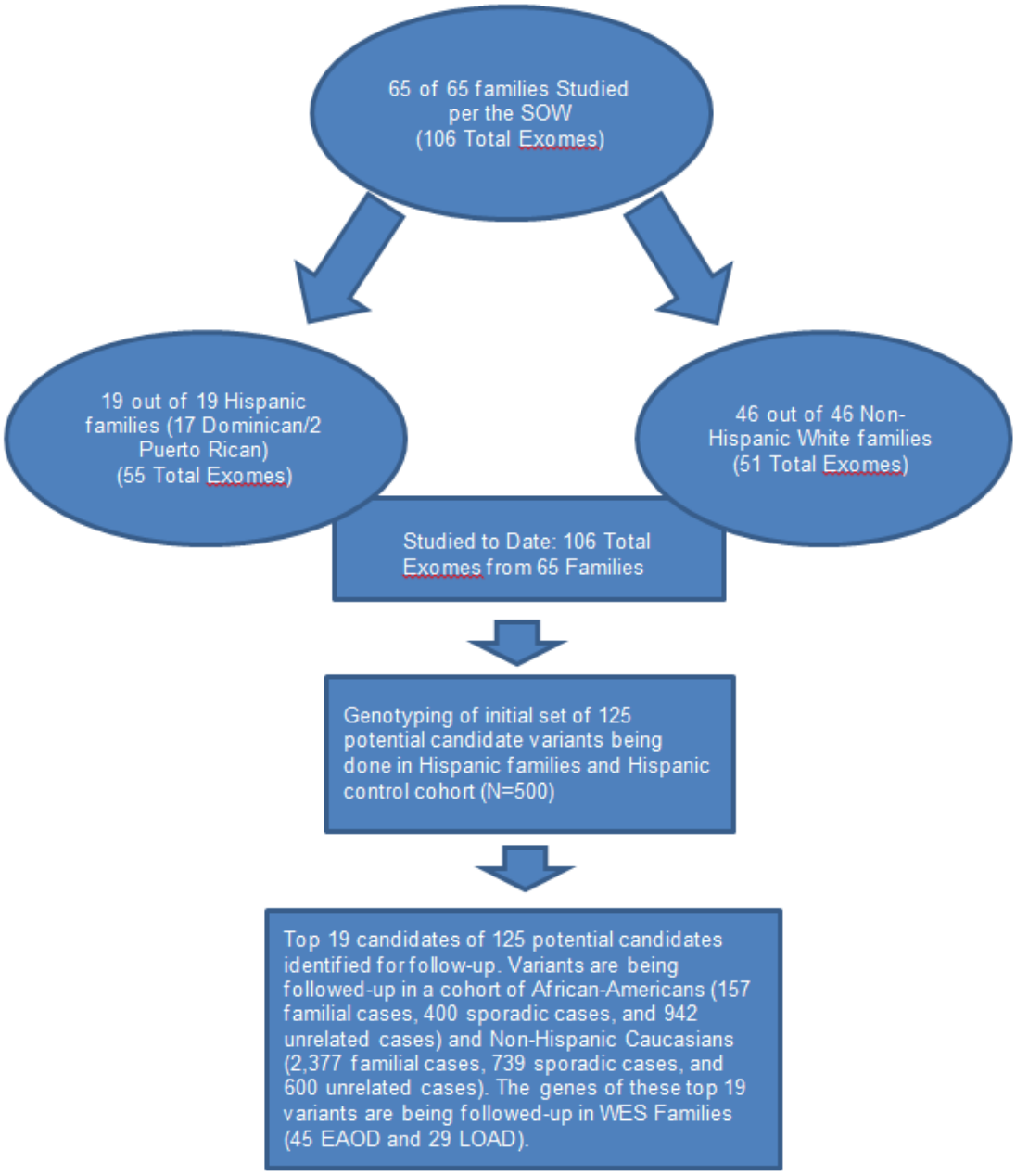
A brief overview of how each family was filtered individually and how variants for typing were prioritized follows:
1) Quality Filter per individual WES sample: VQSLOD>0, PL Score>100, Read Depth>6
2) Annotation of remaining variants with ANNOVAR
3) Remove variants with MAF>0.001 in EVS_6500si and 1000G2012mar_all and MAF>0.01 in HIHG internal controls
4) Keep variants with Autosomal dominant and X-linked dominant segregation in family
4) Exclude variant if not missense, Splicing, Stopgain, Stoploss, Nonframeshift Indel, or Frameshift Indel in refSeq gene annotation, Ensemble gene annotation, or UCSC Known gene annotation
5) Filter on deleteriousness based on a) damaging score in any of these 7 programs programs: Sift, Polyphen2_HDIV, LRT, MutationTaster, MutationAssessor, or FATHMM and b) conservation based on a conserved score in any of these 3 programs: GERP, SiPhy or PhyloP
7) Apply IBD sharing results and require 100% sharing in Hispanic families with enough GWASed individuals
8) Genotype any variant passing above filters and in a known EOAD or LOAD
9) Interrogate shared variants and variants in shared genes across Hispanic Families and between Hispanic and NH-White Families by screening them for existence and potentially too high a MAF in dbSNP, EVS, 1000G updates, specific 1000G populations (EA, AA, AMR and ASN, and any population in UCSC), and cg69 (69 complete genomics exomes). Because of the large amount of candidate genes generated from filtering of the NH-White cases, a variant from the comparison of Hispanic and NH-White candidates was only carried forward for genotyping if the variant/gene passed this screening and was in 2+ Hispanic families and 2+ NH-White cases. Additionally, variants/genes still in 2+ Hispanic families after the screening were carried forward for genotyping.
8) Additional variants were selected by applying a 'secondary filter' to the Hispanic families in order to reduce single variant per family candidates:
---remove any SNV with an rs# in dbSNP129-dbSNP137
---remove all indels
---remove families with greater than 50 variants remaining (families 1,171,386 and 419)
---keep only variants predicted to be damaging in 3 or more of the 7 prediction programs used
---NOTE: Candidate variants for the four removed families were selected based on shared variants/genes with other families.

Follow-up Genotyping of Top Candidate Variants

261 Hispanic familial subjects from 19 pedigrees (145 affecteds and 116 unaffecteds) and 500 Hispanic non-familial subjects (382 healthy controls and 118 sporadic EOAD cases) were genotyped for these 125 top candidate variants. 101 of the variants passed all QC filters (13 variants failed genotyping and 11 were monomorphic in the dataset). For analysis of results of this follow-up genotyping we: 1) estimated familial and population frequencies of the variants in our follow-up cohort and 2) tested single SNV association with AD with 2 models using generalized estimation equations (GEE):
M1) AD~SNV+AGE+SEX
M2) AD~SNV+AGE+SEX+APOE

19 top candidate variants were identified from this follow-up genotyping. They include 8 variants that show perfect segregation with AD status in the families and are absent in population controls. These variants are in the genes *MYO3A, AAAS, DICER1, YIPF1, ACAP1, LLGL2, BPIFB2*, and *ABCG2*. An additional 11 variants were identified as follow-up candidates based on them showing near complete segregation (absent in one or a few familial cases) and being absent in all familial and sporadic controls. These variants are in the genes *GPR26*, *ERCC*6, *OR5M9, DNAH3, MYOCD, KIF17, TICRR, PLXNB2, LAMA2, SNRNP48, and GLB1L2.* These top 19 variants are now being genotyped in a cohort of African-Americans (157 familial cases, 400 sporadic cases, and 942 unrelated cases) and Non-Hispanic Caucasians (2,377 familial cases, 739 sporadic

cases, and 600 unrelated cases). All high priority variants are being genotyped in our large LOAD case control and family based datasets of over 4000 individuals



KEY RESEARCH ACCOMPLISHMENTS:

- Variant calling and quality control processing of these samples completed on 55 Hispanic individuals submitted by Columbia and 51 NH-White samples from the University of Miami and Vanderbilt University.
- Analysis (variant annotation and filtering) completed on samples of 55 Hispanic individuals submitted by Columbia and 51 NH-White samples from the University of Miami and Vanderbilt University.
- Identity-by-descent analysis of Hispanic families is complete.
- Identification of 125 top candidate variants for follow-up genotyping is complete.
- Genotyping of 125 top candidate variants in the Hispanic families and a cohort of 500 Hispanic cases controls is complete.
- Analysis of the 125 top candidate variants in the Hispanic families and a cohort of 500 Hispanic cases and controls is complete, with 19 top candidates identified for follow-up.
- Comparison of the top 19 Hispanic candidates from the follow-up genotyping to the Caucasian EOAD WES samples is ongoing.
- Analysis of candidate variants/loci in our large LOAD case control data set is ongoing.

REPORTABLE OUTCOMES:

Platform Presentation (Appendix I):

Kunkle BW, Kohli MA, Vardarajan BN, Reitz C, Naj AC, Whitehead PL, Martin ER, Beecham GW, Gilbert JR, Farrer LA, Haines JL, Schellenberg GD, Mayeux RP, Pericak-Vance MA, Alzheimer's Disease Genetics Consortium. Whole-exome sequencing in early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer-related genes and processes. The 63rd Annual Meeting of the American Society of Human Genetics (ASHG), Boston, MA, October 22-26, 2013.

Accepted for Platform Presentation (Appendix II):

Reitz C, Kunkle BW, Vandarajan BN, Kohli MA, Naj AC, Whitehead PL, Perry WR, Martin ER, Beecham GW, Gilbert JR, Farrer LA, Haines JL, Schellenberg GD, Pericak-Vance MA, Mayeux RP, Alzheimer's Disease Genetics Consortium. Whole-exome sequencing of Hispanic early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer-related genes. The American Academy of Neurology (AAN) 66th Annual Meeting, Philadelphia, PA, April 26-May 3, 2014.

CONCLUSION:

Mutations in *APP*, *PSEN1* and *PSEN2* lead to familial EOAD and accounting for 60-70% of familial EOAD and ~11% of EOAD overall, leaving the majority of genetic risk for this form of Alzheimer disease unexplained. We performed Whole-Exome Sequencing (WES) on 55 individuals in 19 Caribbean Hispanic EOAD families and 51 Non-Hispanic White EOAD cases previously screened negative for *APP*, *PSEN1* and *PSEN2* to search for rare variants contributing to risk for EOAD. Variants were filtered for segregating, conserved and functional rare variants (MAF<0.1%) assuming both autosomal and X-linked dominant models. 125 rare, segregating, conserved and functional variants passed our stringent filtering criteria for selection of follow-up genotyping candidates. These variants have undergone follow-up genotyping for segregation in the families and for presence in a cohort of 500 Hispanic cases and controls. 19 top candidate variants were identified from this follow-up genotyping. They include 8 variants that show perfect segregation with AD status in the families and are absent in population controls. These variants are in the genes *MYO3A, AAAS, DICER1, YIPF1, ACAP1, LLGL2, BPIFB2*, and *ABCG2*. An additional 11 variants were identified as follow-up candidates based on them showing near complete segregation (absent in one or a few familial cases) and being absent in all familial and sporadic controls. These variants are in the genes *GPR26*, *ERCC*6, *OR5M9, DNAH3, MYOCD, KIF17, TICRR, PLXNB2, LAMA2, SNRNP48, and GLB1L2.* These top 19 variants are now being genotyped in a cohort of African-Americans (157 familial cases, 400 sporadic cases, and 942 unrelated cases) and Non-Hispanic Caucasians (2,377 familial cases, 739 sporadic cases, and 600 unrelated cases). All high priority variants are being genotyped in our large LOAD case control and family based datasets of over 4000 individuals.

APPENDICES:

Appendix I:

Whole-exome sequencing in early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer-related genes and processes

Brian W. Kunkle[1], Martin A. Kohli[1], Badri N. Vardarajan[2], Christiane Reitz[2], Adam C. Naj[3], Patrice L. Whitehead[1], Eden R. Martin[1], Gary W. Beecham[1], John R. Gilbert[1], Lindsay A. Farrer[3], Jonathan L. Haines[4], Gerard D. Schellenberg[5], Richard P. Mayeux[2], Margaret A. Pericak-Vance[1], and The Alzheimer's Disease Genetics Consortium.

[1] John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA
[2] Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY, USA
[3] School of Medicine, Boston University, Boston, MA, USA
[4] Center for Human Genetics Research, Vanderbilt University, Nashville, TN, USA
[5] Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Background

Mutations in *APP*, *PSEN1* and *PSEN2* lead to familial, early-onset Alzheimer disease (EOAD). These mutations account for only 60-70% of familial EOAD and ~11% of EOAD overall, leaving the majority of genetic risk for the most severe form of Alzheimer disease unexplained.

Methods

We performed Whole-Exome Sequencing in Caribbean Hispanic and Caucasian EOAD families previously screened negative for *APP*, *PSEN1*, and *PSEN2* to search for rare variants contributing to risk for EOAD. 60 individuals in 21 families were sequenced using the Agilent 50Mb kit on an Illumina HiSeq2000. Variant filtering for segregating, conserved and functional rare variants (MAF<0.1%) was performed on the 21 families assuming both autosomal-dominant and X-linked dominant models. Filtered loci were examined for implication as AD candidate genes from GWAS or in biologically relevant KEGG Pathways. Variants were also followed up for association with AD in 13,748 individuals (7,652 affected) from the Alzheimer's Disease Genetics Consortium (ADGC) genotyped on the exome chip, which included 195,039 variants with MAF<2%. Enrichment analysis of the variant list was conducted using DAVID.

Results

984 variants in 886 genes passed our stringent filtering criteria, including 63 genes with rare segregating, conserved and functional variants in two or more families. A frameshift mutation in *ABCA7* and a missense variant in *ZCWPW1* are present in one of the 23 GWAS-confirmed Alzheimer disease candidate genes. Seven variants are in AD KEGG Pathway genes (*BID, CYC1, ITPR1, ITPR2, LRP1, ATP2A1*), including two variants in *LRP1,* a gene involved in AD through its roles in cholesterol transport and β-amyloid modulation. Follow up in ADGC exome chip association results comparing EOAD vs. late-onset AD identified 13 of our filtered genes with suggestive associations ($P<10^{-3}$), including *ITM2C* ($P=1.22\times10^{-4}$), a gene known to inhibit the processing of *APP* by blocking access to alpha- and beta-secretase. Enrichment analysis of the list of rare conserved, functional variants showed significant, Benjamini FDR-adjusted enrichment for several AD-related processes including the 'ECM-receptor interaction' and 'ABC transporters' KEGG pathways; GO terms including 'homophilic cell adhesion' and 'microtubule-based movement'; and multiple INTERPRO 'cadherin' classes.

Conclusion

Exome sequencing of EOAD pedigrees identified multiple rare segregating variants with potential roles in AD pathogenesis, several of which were shared in two or more families.

Appendix II:

Whole-exome sequencing of Hispanic early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer-related genes. *C. Reitz[1], B. W. Kunkle[2], B. N. Vardarajan[1], M. A. Kohli[2], A. C. Naj[3], P. L. Whitehead[2], W. R. Perry[2], E. R. Martin[2], G. W. Beecham[2], J. R. Gilbert[2], L. A. Farrer[3], J. L. Haines[4], G. D. Schellenberg[5], M. A. Pericak-Vance[2], R. P. Mayeux[1], Alzheimer's Disease Genetics Consortium* 1) Taub Institute for Research on Alzheimer's Disease, Columbia University, New York, NY, USA; 2) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 4) School of Medicine, Boston University, Boston, MA, USA; 5) Center for Human Genetics Research, Vanderbilt University, Nashville, TN, USA.

OBJECTIVE: To identify novel early-onset Alzheimer disease (EOAD) candidate genes.

BACKGROUND: Mutations in *APP*, *PSEN1* and *PSEN2* lead to familial EOAD and accounting for 60-70% of familial EOAD and ~11% of EOAD overall, leaving the majority of genetic risk for this form of Alzheimer disease unexplained.

DESIGN/METHODS: We performed Whole-Exome Sequencing (WES) on 55 individuals in 19 Caribbean Hispanic EOAD families previously screened negative for *APP*, *PSEN1* and *PSEN2* to search for rare variants contributing to risk for EOAD. Variants were filtered for segregating, conserved and functional rare variants (MAF<0.1%) assuming both autosomal and X-linked dominant models. Filtered loci were examined for implication as AD candidate genes by comparison to: late-onset Alzheimer (LOAD) susceptibility genes, biologically relevant Alzheimer KEGG Pathway genes, candidate genes from 45 WESed NH-White EOAD cases, and results of an Alzheimer's Disease Genetics Consortium (ADGC) exome chip association study.

RESULTS: 2,225 variants in 1,531 genes passed our stringent filtering criteria, including 308 genes with rare segregating, conserved and functional variants in two or more families. Frameshift insertions-deletions in *ABCA7* and *HLA-DRB1*, a nonframeshift deletion in *RIN3*, and missense variants in *DSG2* and *PICALM*, all LOAD susceptibility genes, were discovered. 11 AD KEGG Pathway genes have variants, including *LRP1*, a gene involved in cholesterol transport and β-amyloid modulation. 83 variant carrying genes are in 2+ Hispanic and 2+ Non-white Hispanic families, including the AD-relevant *HLA-A* (associated with earlier age-at-onset), *CHST15* (a potential modulator of Abeta toxicity), and *NOTCH4* (a presenilin pathway gene). Exome chip results identified variants in MICA encoding the *HLA-A* gene and previously associated with LOAD in a small study, as having suggestive association ($p=9.10x10^{-4}$). One family has variants in both *HLA-A* and *MICA*.

CONCLUSIONS: Exome sequencing of Hispanic EOAD pedigrees identified multiple rare segregating variants with potential roles in AD pathogenesis, several of which were shared in two or more families.